

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/938,548A

DATE: 07/24/98
TIME: 12:44:08

INPUT SET: S27673.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

SEQUENCE LISTING

ENTERED

1
2
3 (1) General Information
4
5 (i) APPLICANT: Yanagisawa, Masashi
6 Bergsma, Derk
7 Wilson, Shelagh
8 Brooks, David
9 Gellai, Miklos
10
11 (ii) TITLE OF THE INVENTION:NOVEL LIGANDS OF THE NEUROPEPTIDE
12 RECEPTOR HFGAN72
13
14 (iii) NUMBER OF SEQUENCES:-21--
15
16 (iv) CORRESPONDENCE ADDRESS:
17 (A) ADDRESSEE: SmithKline Beecham Corporation
18 (B) STREET: 709 Swedeland Road
19 (C) CITY: King of Prussia
20 (D) STATE: PA
21 (E) COUNTRY: United States of America
22 (F) ZIP: 19406
23
24 (v) COMPUTER READABLE FORM:
25 (A) MEDIUM TYPE: Diskette
26 (B) COMPUTER: IBM Compatible
27 (C) OPERATING SYSTEM: DOS
28 (D) SOFTWARE: FastSEQ for Windows Version 2.0
29
30 (vi) CURRENT APPLICATION DATA:
31 (A) APPLICATION NUMBER: 08/938,548
32 (B) FILING DATE: 26-SEPT-1997
33 (C) CLASSIFICATION:
34
35 (vii) PRIOR APPLICATION DATA:
36 (A) APPLICATION NUMBER: 08/887,382
37 (B) FILING DATE: 2-JUL-1997
38
39 (vii) PRIOR APPLICATION DATA:
40 (A) APPLICATION NUMBER: 08/820,519
41 (B) FILING DATE: 19-MAR-1997
42
43 (A) APPLICATION NUMBER: 60/033,604
44 (B) FILING DATE: 17-DEC-1997
45
46 (viii) ATTORNEY/AGENT INFORMATION:

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/938,548ADATE: 07/24/98
TIME: 12:44:10

INPUT SET: S27673.raw

47 (A) NAME: King, William T.
48 (B) REGISTRATION NUMBER: 30,954
49 (C) REFERENCE/DOCKET NUMBER: ATG50037-2
50

51 (ix) TELECOMMUNICATION INFORMATION:

52 (A) TELEPHONE: 610-270-5219
53 (B) TELEFAX: 610-270-4026
54 (C) TELEX:
55

56 (2) INFORMATION FOR SEQ ID NO:1:

57 (i) SEQUENCE CHARACTERISTICS:

58 (A) LENGTH: 1970 base pairs
59 (B) TYPE: nucleic acid
60 (C) STRANDEDNESS: single
61 (D) TOPOLOGY: linear
62

63 (ii) MOLECULE TYPE: Genomic DNA

64 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

65
66
67
68 AAAACATAAT GTGGGTCTCG CGTCTGCCTC TCTCCCGCCC CTAATTAGCA GCTGCCTCCC 60
69 TCCATATTGT CCCAGGCCAG CGCTTCTTTT GTGCTCCAG ATTCTTGGGT GCAAGGTGGC 120
70 CTCATTAGTG CCCGGAGACC GCCCCATCTC CAGGGAGCAG ATAGACAGAC AAGGGGGTGA 180
71 TCAGGGGCAC AGTGATCCAA CCCTGGCCTC TGAACGCCGC AGCGGCCATT CCTTGGGCCC 240
72 AGCCTGGAGA CGGCCCCCCT GCAGCAGGCT AATCTTAGAC TTGCCTTTGT CTGGCCTGGG 300
73 TGTGGACGCA ATGTGCCTGT CAATTCCCCG CCACCTCAGA GCACTATAAA CCCAGACCC 360
74 CTGGGAGTGG GTCACAATTG ACAGCCTCAA GGTTCCCTGGC TTTTGAACC ACCACAGACA 420
75 TCTCCTTTCC CGGTACCCC ACCCTGAGCG CCAGACACCA TGAACCTTCC TTCCACAAAG 480
76 GTAAAGATCC AGGGATGGAG GGGTGACTCA GCCATCCCAG AGGAAGCAAA AAGAGTGCTT 540
77 GCTCAGAGGG CTGGAAGAAA GGCCAAAGGT GTCTCCACTC TTGGTCTTTT CCTGGGTGTG 600
78 CTCTGAGGCA GGAGCACCTG CCTTGGCTCA CATTGGGTGT GGTGCTGTTT TGCTAAGAGC 660
79 CTGTGTTTGC TGAGCTCATA TGTGTCAGGT GCTCCGTTTG CACCTGTCAT CTCTTGTCAT 720
80 CCTCCAACA GCCTTGCAGA GTAGAAATTA TTTCTAGTAT ACCCAGTTTA CAGGTAAGGG 780
81 AGCTGTGCCC TCTGAAAGGG CAGGAACTG GTTCAAAGCA ACGGAGTTCA GTCACCTCTG 840
82 CAAGGGGGCA GGCAGATGAG AGAGCATTC TGGAGTCTTG TAGTTCCTGA TTTCCATGTG 900
83 TTTCCCTGCT GTGGAGAGGA AGTTGGGGGG ACTCAGTAGG GCCCGGGTTT TTCCCAAGTT 960
84 TACAACCTCT GCTGCAGACA GACACTCCTG TTTTCAGGTG GAGTGGCAAG TGCCCTAGTG 1020
85 GTGGCAACAG TGGCCTAAGT CTCCAGAGAA AAGGGGGATT CACTCTGCCC AGGGGTCTC 1080
86 AAAAGGCTTC CTGTGGGAGA TGCTCTGCTG GGTCTTGAAG GAGGAGCAGG GAAAGTAGGC 1140
87 CGATACCAGC AAGGGCGCAA AGCAAGGAGA ACTAAGTGAC AGCCAGAAAG GAGTGCAGGC 1200
88 TTGGAGGGGG CGCGGAGCCA GAGGGGCAGG TCCTGTGCGT GGGAGCTGGT GCGGGGCGCC 1260
89 GTGGGAAGAC CCCCCAGCG CCCTGTCTCC GTCTCCCTAG GTCTCCTGGG CCGCCGTGAC 1320
90 GCTACTGCTG CTGCTGCTGC TGCTGCCGCC CGCGCTGTTG TCGTCCGGGG CGGCTGCACA 1380
91 GCCCCTGCCC GACTGCTGTC GTCAAAAGAC TTGCTCTTGC CGCCTCTACG AGCTGCTGCA 1440
92 CGGCGCGGGC AATCACGCGG CCGGCATCCT CACGCTGGGC AAGCGGAGGT CCGGGCCCCC 1500
93 GGGCCTCCAG GGTCCGCTGC AGCGCCTCCT GCAGGCCAGC GGCAACCACG CCGCGGGCAT 1560
94 CCTGACCATG GGCCGCCGCG CAGGCGCAGA GCCAGCGCCG CGCCCCTGCC TCGGGCGCCG 1620
95 CTGTTCCGCC CCGGCCGCCG CCTCCGTCGC GCCCGGAGGA CAGTCCGGGA TCTGAGTCGT 1680
96 TCTTCGGGCC CTGTCTTGGC CCAGGCCTCT GCCCTCTGCG CACCCAGCGT CAGCCCCAG 1740
97 AAAAAAGGCA ATAAAGACGA GTCTCCATTC GTGTGACTGG TCTCTGTTCC TGTGCGGTCTG 1800
98 CGTCTGCCC ATCCGGGGTG GCAAAGCGTC TTGCGGAGGA CAGCTGGGCC TGGAAGCCCG 1860
99 GCTGTCGGGC ACCAGCCTTA GCTTTTGCGT GGTGAATCG GAAACACTCT TGGTTGGGGA 1920

RAW SEQUENCE LISTING PATENT APPLICATION US/08/938,548A

DATE: 07/24/98
TIME: 12:44:11

INPUT SET: S27673.raw

100 GTTCCCAGTG CAAGGCCCTG GGGCACAGAG AGAACTGCAC AGGTGCATGC 1970

101

102 (2) INFORMATION FOR SEQ ID NO:2:

103

104 (i) SEQUENCE CHARACTERISTICS:

105 (A) LENGTH: 131 amino acids

106 (B) TYPE: amino acid

107 (C) STRANDEDNESS: single

108 (D) TOPOLOGY: linear

109

110 (ii) MOLECULE TYPE: protein

111

112 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

113

114 Met Asn Leu Pro Ser Thr Lys Val Ser Trp Ala Ala Val Thr Leu Leu

115 1 5 10 15

116 Leu Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Ser Gly Ala Ala

117 20 25 30

118 Ala Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg

119 35 40 45

120 Leu Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu

121 50 55 60

122 Thr Leu Gly Lys Arg Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu

123 65 70 75 80

124 Gln Arg Leu Leu Gln Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr

125 85 90 95

126 Met Gly Arg Arg Ala Gly Ala Glu Pro Ala Pro Arg Pro Cys Leu Gly

127 100 105 110

128 Arg Arg Cys Ser Ala Pro Ala Ala Ser Val Ala Pro Gly Gly Gln

129 115 120 125

130 Ser Gly Ile

131 130

132

133 (2) INFORMATION FOR SEQ ID NO:3:

134

135 (i) SEQUENCE CHARACTERISTICS:

136 (A) LENGTH: 33 amino acids

137 (B) TYPE: amino acid

138 (C) STRANDEDNESS: single

139 (D) TOPOLOGY: linear

140

141 (ii) MOLECULE TYPE: protein

142

143 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

144

145 Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu

146 1 5 10 15

147 Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr

148 20 25 30

149 Leu

150

151 (2) INFORMATION FOR SEQ ID NO:4:

152

RAW SEQUENCE LISTING PATENT APPLICATION US/08/938,548A

DATE: 07/24/98
TIME: 12:44:13

INPUT SET: S27673.raw

153 (i) SEQUENCE CHARACTERISTICS:
154 (A) LENGTH: 28 amino acids
155 (B) TYPE: amino acid
156 (C) STRANDEDNESS: single
157 (D) TOPOLOGY: linear
158
159 (ii) MOLECULE TYPE: protein
160
161 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
162
163 Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln
164 1 5 10 15
165 Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr Met
166 20 25
167

168 (2) INFORMATION FOR SEQ ID NO:5:
169

170 (i) SEQUENCE CHARACTERISTICS:
171 (A) LENGTH: 585 base pairs
172 (B) TYPE: nucleic acid
173 (C) STRANDEDNESS: single
174 (D) TOPOLOGY: linear
175

176 (ii) MOLECULE TYPE: cDNA
177

178 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
179

180	GGCTCGGCGG	CCTCAGACTC	CTTGGGTATT	TGGACCACTG	CACCGAAGAT	ACCATCTCTC	60
181	CGGATTGCCT	CTCCCTGAGC	TCCAGACACC	ATGAACCTTC	CTTCTACAAA	GGTTCCCTGG	120
182	GCCGCCGTGA	CGCTGCTGCT	GCTGCTACTG	CTGCCGCCGG	CGCTGCTGTC	GCTTGGGGTG	180
183	GACGCGCAGC	CTCTGCCCGA	CTGCTGTTCG	CAGAAGACGT	GTTCCTGCCG	TCTCTACGAA	240
184	CTGTTGCACG	GAGCTGGCAA	CCACGCCGCG	GGCATCCTCA	CTCTGGGAAA	GCGGCGACCT	300
185	GGACCCCCAG	GCCTCCAAGG	ACGGCTGCAG	CGCCTCCTTC	AGGCCAACGG	TAACCACGCA	360
186	GCTGGCATCC	TGACCATGGG	CCGCCGCGCA	GGCGCAGAGC	TAGAGCCATA	TCCCTGCCCT	420
187	GGTCGCCGCT	GTCCGACTGC	AACCGCCACC	GCTTTAGCGC	CCCGGGGCGG	ATCCAGAGTC	480
188	TGAACCCGTC	TTCTATCCCT	GTCCTAGTCC	TAACTTTCCC	CTCTCCTCGC	CGGTCCCTAG	540
189	GCAATAAAGA	CGTTTCTCTG	CTAAAAAAA	AAAAAAAAAA	AAAAA		585

190
191 (2) INFORMATION FOR SEQ ID NO:6:
192

193 (i) SEQUENCE CHARACTERISTICS:
194 (A) LENGTH: 130 amino acids
195 (B) TYPE: amino acid
196 (C) STRANDEDNESS: single
197 (D) TOPOLOGY: linear
198

199 (ii) MOLECULE TYPE: protein
200

201 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
202

203 Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
204 1 5 10 15
205 Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala

RAW SEQUENCE LISTING PATENT APPLICATION US/08/938,548A

DATE: 07/24/98
TIME: 12:44:15

INPUT SET: S27673.raw

```

206          20          25          30
207  Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
208          35          40          45
209  Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
210          50          55          60
211  Leu Gly Lys Arg Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln
212  65          70          75          80
213  Arg Leu Leu Gln Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
214          85          90          95
215  Gly Arg Arg Ala Gly Ala Glu Leu Glu Pro Tyr Pro Cys Pro Gly Arg
216          100          105          110
217  Arg Cys Pro Thr Ala Thr Ala Thr Ala Leu Ala Pro Arg Gly Gly Ser
218          115          120          125
219  Arg Val
220          130

```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

234  Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
235  1          5          10          15
236  Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
237          20          25          30

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

251  Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
252  1          5          10          15
253  Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
254          20          25          30
255  Leu

```

(2) INFORMATION FOR SEQ ID NO:9:

PAGE: 1

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION US/08/938,548A

DATE: 07/24/98
TIME: 12:44:17

INPUT SET: S27673.raw

Line

Error

Original Text